Journal of Computational and Applied Mathematics 223 (2009) 1013–1027

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

A method combining norm-relaxed QP subproblems with systems of linear equations for constrained optimization[☆]

Jin-bao Jian[a,1], Xiao-yan Ke[b], Hai-yan Zheng[a], Chun-ming Tang[a,*]

[a] College of Mathematics and Information Science, Guangxi University, Nanning, Guangxi 530004, China

[b] China Liaohe Petroleum Engineering Co., Ltd, Liaoning, China

ARTICLE INFO

Article history:
Received 30 October 2007
Received in revised form 21 March 2008

MSC:
90C30
49M37
65K05

Keywords:
Constrained optimization
Norm-relaxed SQP method
Systems of linear equations
Method of strongly sub-feasible directions
Global convergence
Superlinear convergence

ABSTRACT

Based on the ideas of norm-relaxed sequential quadratic programming (SQP) method and the strongly sub-feasible direction method, we propose a new SQP algorithm for the solution of nonlinear inequality constrained optimization. Unlike the previous work, at each iteration, the norm-relaxed quadratic programming subproblem (NRQPS) in our algorithm only consists of the constraints corresponding to an estimate of the active set, and the high-order correction direction (used to avoid the Maratos effect) is obtained by solving a system of linear equations (SLE) which also only consists of such a subset of constraints and gradients. Moreover, the line search technique can effectively combine the initialization process with the optimization process, and therefore (if the starting point is not feasible) the iteration points always get into the feasible set after a finite number of iterations. The global convergence is proved under the Mangasarian–Fromovitz constraint qualification (MFCQ), and the superlinear convergence is obtained without assuming the strict complementarity. Finally, the numerical experiments show that the proposed algorithm is effective and promising for the test problems.

1. Introduction

Consider the following nonlinear inequality constrained optimization problem:

$$\begin{array}{ll} \min\limits_{x} & f(x) \\ \text{s.t.} & g_j(x) \le 0, \quad j \in I = \{1, 2, \ldots, m\}, \end{array} \tag{1.1}$$

where functions $f, g_j\ (j \in I) : R^n \to R$ are all continuously differentiable.

As is well known, the sequential quadratic programming (SQP) algorithm is one of the most effective algorithms available for solving problem (1.1). However, in the traditional SQP methods, there are two disadvantages: the quadratic programming (QP) subproblems may be inconsistent and the Maratos effect [1] may occur. One way to overcome the former disadvantage is to generate iterates lying within the feasible set of (1.1), and this leads to a class of feasible SQP algorithms (see e.g. [2,3, 5–18]). A popular way to the latter is to use a high-order direction generated by solving a QP subproblem [2,3] or a system of linear equations [4], or directly by an explicit formulas [5–9].

☆ Project supported by the National Natural Science Foundation (No. 10771040) and Guangxi Science Foundation (No. 0640001) of China as well as Guangxi University Key Program for Science and Technology Research (No. 2005ZD02).

* Corresponding author.
E-mail addresses: jianjb@gxu.edu.cn (J.-b. Jian), ronger6666@163.com (X.-y. Ke), zhy166@gxu.edu.cn (H.-y. Zheng), cmtang@gxu.edu.cn (C.-m. Tang).

1 Fax: +86 7713237018.

0377-0427/$ – see front matter © 2008 Published by Elsevier B.V.
doi:10.1016/j.cam.2008.03.048

The norm-relaxed method of feasible directions (MFD) is an important kind of feasible SQP method, which was proposed in [10], and further improved in [11]. In particular, the direction finding subproblem (DFS) in [11] has the following form:

$$\begin{aligned} \min_{(z,d)} \quad & z + \frac{1}{2} d^T B_k d \\ \text{s.t.} \quad & \nabla f(x^k)^T d \le \gamma_0 z, \\ & g_j(x^k) + \nabla g_j(x^k)^T d \le \gamma_j z, \quad j \in I, \end{aligned}$$

where $z \in R$ is an additional variable, B_k is a positive definite matrix and γ_0, γ_j $(j \in I)$ are all positive constants. In order to obtain superlinear convergence, this kind of method was further studied, see e.g., Kostreva and Chen [12], Lawrence and Tits [13] as well as Zhu and Zhang [14]. In order to overcome the Maratos effect, a high-order correction direction is generated by solving another QP subproblem [12,13] or by an explicit formula [14]. We note that in the above three algorithms [12–14] the strict complementarity assumption is required, which is relatively strong and difficult for checking. Moreover, these algorithms have a common disadvantage that the initial point must be feasible, and therefore an initial feasible point must be obtained in advance, but this is usually a nontrivial work.

Recently, based on the ideas of the strongly sub-feasible direction method [15] and norm-relaxed MFD method, Jian et al. [16] proposed a globally convergent norm-relaxed method of strongly subfeasible directions. This algorithm can not only combine the initialization process (to find a feasible starting point) with the optimization process (to get an optimal solution), but also guarantee that the number of satisfied constraints (i.e., $g_j(x) \le 0$) is nondecreasing. However, the algorithm in [16] is only globally convergent, since no high-order correction direction is used. More recently, the algorithm in [16] is further improved by the authors [17], in which the DFS subproblem has the following form:

$$\begin{aligned} \min_{(z,d)} \quad & \gamma_0 z + \frac{1}{2} d^T B_k d \\ \text{s.t.} \quad & \nabla f(x^k)^T d \le \gamma_0 z + \gamma \varphi(x^k)^\sigma, \\ & g_j(x^k) + \nabla g_j(x^k)^T d \le \gamma_j \eta_k z, \quad j \in I^-(x^k), \\ & g_j(x^k) + \nabla g_j(x^k)^T d \le \gamma_j \eta_k z + \varphi(x^k), \quad j \in I^+(x^k), \end{aligned}$$

where $I^-(x^k) = \{j \in I : g_j(x^k) \le 0\}, I^+(x^k) = \{j \in I : g_j(x^k) > 0\}, \eta_k$ is a positive parameter associated with $x^k, \gamma, \sigma, \gamma_j$ $(j = 0, 1, \ldots, m)$ are all positive constant parameters, $\varphi(x^k)$ is the maximal violated constraints function. They used a high-order direction which is generated by an explicit formula to avoid the Maratos effect, and then the superlinear convergence is proved under mild conditions without the strict complementarity. However, it can be seen that the constraints of the DFS above consists of all the indices corresponding to the constraints of the original problem (1.1), thus the scale of the DFS subproblem is large if the number of constraints m is large. Furthermore, the high-order updated direction used in [17] also includes all the constraints corresponding to the original problem (1.1), and an inverse matrix must be computed. In addition, before the iteration points getting into the feasible set, the arc search ignores the variation of the objective function value, as a result, the iteration point generated by the initialization phase may deviate the solution of the problem (1.1).

Building on the observations above, in this paper, we aim to improve the algorithm in [17]. At first, we introduce an active set technique to remove those constraints that are locally irrelevant, and therefore the scale of the DFS subproblem is largely decreased. Secondly, instead of an explicit formula, we generate a high-order direction by solving a system of linear equations which also only includes the estimate of the active constraints and their gradients, and this is generally more efficient than computing an inverse matrix. At last, the line search used in the new algorithm can well combine the initialization and optimization processes, and meanwhile prevent the objective value from increasing too much.

The main features of the proposed algorithm are summarized as follows:

- if the algorithm starts from an infeasible point, the number of satisfied constraints is nondecreasing. Moreover, after a finite number of iterations, the iteration points always get into the feasible set;
- the improved direction and the high-order updated direction are obtained by solving one sub-problem and a system of linear equations, respectively. Both them only depend on the constraints and their gradients corresponding to an estimate of the active set;
- the global convergence is proved under weak conditions (e.g., MFCQ), and the superlinear convergence is obtained without assuming the strict complementarity.

The paper is organized as follows. In the next section, we present the details of our algorithm and discuss its properties. In Section 3, we analyze its global convergence. Strong and superlinear convergence is proved in Section 4. Some preliminary numerical results are reported in Section 5.

2. Description of algorithm

Denote the feasible set for problem (1.1) by $X = \{x \in R^n : g_j(x) \le 0, j \in I\}$. For $x \in R^n$ and parameter $\varepsilon \ge 0$, we use the following notions:

$$I^-(x) = \{j \in I : g_j(x) \le 0\}, \qquad I^+(x) = \{j \in I : g_j(x) > 0\}, \qquad \varphi(x) = \max\{0; g_j(x), j \in I\}, \tag{2.1}$$

$$\left.\begin{array}{l} I^+(x,\varepsilon) = \{j \in I^+(x) : -\varepsilon \le g_j(x) - \varphi(x) \le 0\}, \\ I^-(x,\varepsilon) = \{j \in I^-(x) : -\varepsilon \le g_j(x) \le 0\}, \qquad I(x,\varepsilon) = I^-(x,\varepsilon) \cup I^+(x,\varepsilon). \end{array}\right\} \tag{2.2}$$

Before giving the algorithm, we suppose that the following assumption holds in this paper.

Assumption A1. Functions f, g_j $(j \in I)$ are all continuously differentiable.

For a given iteration point $x^k \in R^n$ and a symmetric positive definite matrix B_k, letting $I_k = I(x^k, \varepsilon)$, we consider the following subproblem:

$$\text{DFS} \quad \begin{array}{ll} \min\limits_{(z,d)} & \gamma_0 z + \dfrac{1}{2} d^T B_k d \\ \text{s.t.} & \nabla f(x^k)^T d \le \gamma_0 z + \gamma \varphi(x^k)^\sigma, \\ & g_j(x^k) + \nabla g_j(x^k)^T d \le \gamma_j \eta_k z, \quad j \in I_k^- \triangleq I^-(x^k, \varepsilon), \\ & g_j(x^k) + \nabla g_j(x^k)^T d \le \gamma_j \eta_k z + \varphi(x^k), \quad j \in I_k^+ \triangleq I^+(x^k, \varepsilon), \end{array} \tag{2.3}$$

where η_k is a positive parameter associated with x^k, and γ, σ, γ_j $(j \in \{0\} \cup I_k)$ are all positive constants, and $I_k = I_k^- \cup I_k^+$.

We will show that DFS (2.3) is well defined under the following constraint qualification that also plays an important role in ensuring the global convergence of the algorithm.

Assumption A2. Suppose that the Mangasarian–Fromovitz constraint qualification (MFCQ) is satisfied by (1.1) at any $x \in R^n$, i.e., there exists a vector $d \in R^n$ such that $\nabla g_j(x)^T d < 0$, $\forall j \in I(x, 0)$.

The MFCQ above is also used in [16], and it is weaker than the linear independence constraint qualification (LICQ) used in many literatures, see e.g. [2,3,5–7,12,14,17].

The lemma given below describes some nice properties of DFS (2.3).

Lemma 2.1. *Suppose that Assumptions A1 and A2 hold, the matrix B_k is symmetric positive definite, and the parameters η_k, γ_j $(j \in \{0\} \cup I_k)$, σ as well as γ are all positive. Then*

(i) *DFS (2.3) has a unique solution;*
(ii) *(z_k, d^k) is an optimal solution of (2.3) if and only if it is a KKT point for (2.3);*
(iii) *$\gamma_0 z_k + \frac{1}{2}(d^k)^T B_k d^k \le 0$, $z_k \le 0$;*
(iv) *$z_k = 0 \Longleftrightarrow d^k = 0 \Longleftrightarrow x^k$ is a KKT point for (1.1).*

The proof of this lemma is similar to that in [16].

Remark 1. From Lemma 2.1, if $x^k \in X$ but x^k is not a KKT point for problem (1.1), then we have $z_k < 0$, and therefore the results that $\nabla f(x^k)^T d^k \le \gamma_0 z_k < 0$ and $\nabla g_j(x^k)^T d^k \le \gamma_j \eta_k z_k < 0$ for $j \in I(x^k, 0) = I^-(x^k, 0)$ hold. So we can conclude that d^k is a feasible direction of descent for problem (1.1) at point x^k. On the other hand, if $x^k \notin X$, it is not necessary to make the objective value of problem (1.1) decrease.

Suppose that (z_k, d^k) is a solution of (2.3), so it is also a KKT point for (2.3) and there is a corresponding multiplier vector $(\mu_k, \lambda_{I_k}^k)$ with $\lambda_{I_k}^k = (\lambda_j^k, j \in I_k)$ such that

$$\gamma_0 \mu_k + \eta_k \sum_{j \in I_k} \gamma_j \lambda_j^k = \gamma_0, \tag{2.4}$$

$$B_k d^k + \mu_k \nabla f(x^k) + \sum_{j \in I_k} \lambda_j^k \nabla g_j(x^k) = 0, \tag{2.5}$$

$$0 \le \mu_k \perp (-\nabla f(x^k)^T d^k + \gamma_0 z_k + \gamma \varphi(x^k)^\sigma) \ge 0, \tag{2.6}$$

$$0 \le \lambda_j^k \perp (-g_j(x^k) - \nabla g_j(x^k)^T d^k + \gamma_j \eta_k z_k) \ge 0, \quad j \in I_k^-, \tag{2.7}$$

$$0 \le \lambda_j^k \perp (-g_j(x^k) - \nabla g_j(x^k)^T d^k + \gamma_j \eta_k z_k + \varphi(x^k)) \ge 0, \quad j \in I_k^+, \tag{2.8}$$

where the notation $x \perp y$ means $x^T y = 0$.

In order to overcome the Maratos effect, a suitable high-order correction direction of d^k is often adopted. In this paper, we introduce a new system of linear equations in (d, h) as follows to generate such a direction

$$V_k \begin{pmatrix} d \\ h \end{pmatrix} = \begin{pmatrix} B_k & N_k \\ N_k^T & -D^k \end{pmatrix} \begin{pmatrix} d \\ h \end{pmatrix} = \begin{pmatrix} 0 \\ -\max\{\|d^k\|^\tau, |\eta_k^\nu z_k| \|d^k\|\} e_{I_k} - \tilde{g}^k \end{pmatrix}, \tag{2.9}$$

where $e_{I_k} = (1, \ldots, 1)^T \in R^{|I_k|}$, $\tau \in (2, 3)$, $\nu \in (0, 1)$, $p > 0$ and

$$V_k \triangleq \begin{pmatrix} B_k & N_k \\ N_k^T & -D^k \end{pmatrix}, \quad N_k = (\nabla g_j(x^k), j \in I_k), D^k = \mathrm{diag}(D_j^k, j \in I_k), \tilde{g}^k = (\tilde{g}_j^k, j \in I_k), \tag{2.10}$$

$$\tilde{g}_j^k = \begin{cases} g_j(x^k + d^k) - g_j(x^k) - \nabla g_j(x^k)^T d^k + \gamma_j \eta_k z_k + \varphi(x^k)^\sigma, & \text{if } j \in I_k^-; \\ g_j(x^k + d^k) - g_j(x^k) - \nabla g_j(x^k)^T d^k + \gamma_j \eta_k z_k + \varphi(x^k) + \varphi(x^k)^\sigma, & \text{if } j \in I_k^+, \end{cases} \tag{2.11}$$

$$D_j^k = \begin{cases} |g_j(x^k)|^p \left(|g_j(x^k) + \nabla g_j(x^k)^T d^k - \gamma_j \eta_k z_k| + |\eta_k z_k| + \|d^k\|\right), & \text{if } j \in I_k^-; \\ |\varphi(x^k) - g_j(x^k)|^p. \\ (|g_j(x^k) + \nabla g_j(x^k)^T d^k - \gamma_j \eta_k z_k - \varphi(x^k)| + |\eta_k z_k| + \|d^k\|), & \text{if } j \in I_k^+. \end{cases} \tag{2.12}$$

The following lemma shows that the system of linear equations (2.9) is well defined.

Lemma 2.2. *Suppose that Assumption A1 holds. If the LICQ holds at x^k, i.e., the vectors $\{\nabla g_j(x^k), j \in I(x^k, 0)\}$ are linearly independent, then the coefficient matrix V_k is nonsingular and (2.9) has a unique solution.*

Proof. Let $u \in R^n$ and $v \in R^{|I_k|}$ be a solution of

$$\begin{pmatrix} B_k & N_k \\ N_k^T & -D^k \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = 0. \tag{2.13}$$

It is sufficient to verify (u, v) must equal zero. Multiplying the first equation by u^T and the second equation by v^T both from left sides, we have

$$u^T B_k u + u^T N_k v = 0, \qquad v^T N_k^T u - v^T D^k v = 0,$$

furthermore,

$$u^T B_k u + v^T D^k v = 0.$$

Since B_k is positive definite and the matrix D^k is diagonal and positive semidefinite, we have

$$u = 0, \qquad D_j^k v_j^2 = 0, \quad \forall j \in I_k.$$

In view of the definition of D^k (2.12), we know that

$$D_j^k \geq 0, \quad \forall j \in I_k, \qquad D_j^k > 0, \quad \forall j \in I_k \setminus I(x^k, 0),$$

so $v_j = 0, \forall j \in I_k \setminus I(x^k, 0)$. This together with the first equation of (2.13) gives

$$0 = N_k v = \sum_{j \in I(x^k, 0)} v_j \nabla g_j(x^k).$$

So the formulas above together with the LICQ show that $v = 0$. Hence V_k is nonsingular and thus (2.9) has a unique solution. □

Based on the properties above, we can present the algorithm as follows.

Algorithm

Step 0. Choose parameters $\varepsilon > 0, c_0 > 0, p > 0, \tau \in (2, 3), \beta \in (0, 1), \alpha \in (0, 0.5), 0 < \sigma' < \sigma \leq 1, \nu \in (0, 1), \gamma, \gamma_j > 0, j = 0, 1, \ldots, m, \xi, \zeta > 0, \eta_0 > 0$, a starting point $x^0 \in R^n$, and an initial symmetric positive definite matrix B_0; set $k := 0$.

Step 1. Solve DFS (2.3) to obtain an optimal solution (z_k, d^k) with multiplier $(\mu_k, \lambda_{I_k}^k)$. If $d^k = 0$, stop; otherwise go to Step 2.

Step 2. Obtain high-order direction $\tilde{d}^k$ by solving the linear system (2.9) if it has a solution. If (2.9) has no solution or $\|\tilde{d}^k\| > \|d^k\| + \varphi(x^k)^{\sigma'}$, then set $\tilde{d}^k = 0$.

Step 3. If $\varphi(x^k) > 0$ and $g_j(x^k + d^k + \tilde{d}^k) \leq 0$ for any $j \in I$, then let $t_k = 1$ and go to step 4, otherwise compute the step size t_k which is the first value of t in the sequence $\{1, \beta, \beta^2, \beta^3, \ldots\}$ that satisfies the following inequalities:

$$f(x^k + td^k + t^2\tilde{d}^k) \leq f(x^k) + \alpha t \nabla f(x^k)^T d^k + (1 - \alpha) t \gamma \varphi(x^k)^\sigma, \tag{2.14}$$

$$g_j(x^k + td^k + t^2\tilde{d}^k) \leq \varphi(x^k) + \frac{t}{2} \gamma_j \eta_k z_k, \quad j \in I^+(x^k), \tag{2.15}$$

$$g_j(x^k + td^k + t^2\tilde{d}^k) \leq 0, \quad j \in I^-(x^k). \tag{2.16}$$

Step 4. Compute η_{k+1} by $\eta_{k+1} = \min\{c_0, \|d^k\|^\xi + |z_k|^\zeta + \varphi(x^k)\}$, and a new symmetric positive definite matrix B_{k+1} by some suitable techniques. Set $x^{k+1} = x^k + t_k d^k + t_k^2 \tilde{d}^k$, $k := k + 1$, and go back to Step 1.

Remark 2. The inequality (2.14) can prevent the objective function from increasing too much when $x^k \notin X$. However, (2.14) can be removed when x^k is not a feasible point, and it has no influence on all the theoretical analysis and may decrease the computational cost.

The following lemma shows that Step 3 can be terminated after a finite number of computations, and thus the algorithm is well-defined.

Lemma 2.3. *The arc search at Step* 3 *can be carried out, that is, the inequalities* (2.14)–(2.16) *hold for* $t > 0$ *small enough.*

Proof. In view of $d^k \neq 0$ and Lemma 2.1, we know that $z_k < 0$, so the rest of the proof is similar to that in [17], and it is omitted. □

Based on the arc search (2.14)–(2.16), we know that exactly one of the following two cases takes place:
Case A: There exists an iteration index s such that $\varphi(x^s) = 0$. In this case, one knows that $\varphi(x^k) = 0$ for all $k \geq s$;
Case B: For any $k = 0, 1, 2, \ldots, \varphi(x^k) > 0, \varphi(x^{k+1}) < \varphi(x^k)$ and $I^-(x^{k+1}) \supseteq I^-(x^k)$.

3. Convergence analysis

If the algorithm stops at x^k, we know from Lemma 2.1 and Step 1 that x^k is a KKT point for (1.1). In this section, assuming that the algorithm generates an infinite iteration sequence $\{x^k\}$ of points, will prove that any accumulation point x^* of $\{x^k\}$ is a KKT point for (1.1). For this purpose, following assumption is necessary.

Assumption A3. The sequence $\{B_k\}$ of matrices is uniformly positive definite, i.e., there exist two positive constants a and b such that

$$a\|d\|^2 \leq d^T B_k d \leq b\|d\|^2, \quad \forall d \in R^n, \forall k.$$

For simplicity, we denote the active set of (2.3) by

$$\left.\begin{array}{l} L_k^+ = \{j \in I_k^+ : g_j(x^k) + \nabla g_j(x^k)^T d^k = \gamma_j \eta_k z_k + \varphi(x^k)\}; \\ L_k^- = \{j \in I_k^- : g_j(x^k) + \nabla g_j(x^k)^T d^k = \gamma_j \eta_k z_k\}, \qquad L_k = L_k^+ \cup L_k^-. \end{array}\right\} \tag{3.1}$$

Suppose that x^* is a given accumulation point of $\{x^k\}$. Then, in view of I_k^+, I_k^-, L_k^+ and L_k^- all being the subsets of the fixed and finite set I, we can assume without loss of generality that there exists an infinite index set K such that

$$\left.\begin{array}{lllll} x^k \to x^*, & B_k \to B_*, & \eta_k \to \eta_*, & \varphi(x^k) \to \varphi(x^*), & k \in K, \\ I_k^- \equiv \tilde{I}^-, & I_k^+ \equiv \tilde{I}^+, & I_k \equiv \tilde{I}^+ \cup \tilde{I}^- \triangleq \tilde{I}, & \forall k \in K, & \\ L_k^+ \equiv L^+, & L_k^- \equiv L^-, & L_k \equiv L^+ \cup L^- \triangleq L, & \forall k \in K. & \end{array}\right\} \tag{3.2}$$

Lemma 3.1. *If Assumptions* A1–A3 *hold, then*

(i) *the sequences* $\{z_k : k \in K\}$, $\{d^k : k \in K\}$ *and* $\{\tilde{d}^k : k \in K\}$ *are all bounded;*
(ii) *the KKT multiplier sequences* $\{\mu_k\}$ *and* $\{\lambda^k \triangleq (\lambda_{\tilde{I}}^k, 0_{I\setminus\tilde{I}}) : k \in K\}$ *of DFS are all bounded;*
(iii) *there exists a constant* c *such that* $\|V_k^{-1}\| \leq c$ *for* $k \in K$ *large enough if the LICQ holds the limit point* x^*.

Proof. Under A1, A3 and the MFCQ (rather than the LICQ used in [17]), the proof of (i) and (ii) similar to that in [17].
(iii) Suppose by contradiction that there exists an infinite index set $K' \subseteq K$ such that

$$\|V_k^{-1}\| \to \infty, \quad K' \subseteq K. \tag{3.3}$$

In view of Assumption A3 and (i), we can assume without loss of generality that there exists $K'' \subseteq K' \subseteq K$ such that

$$B_k \to B_*, \qquad V_k \to V_* = \begin{pmatrix} B_* & N_* \\ N_*^T & -D^* \end{pmatrix}, \qquad d^k \to d^*, \qquad z_k \to z_*, \qquad \eta_k \to \eta_* \qquad I_k \equiv \tilde{I}, \quad k \in K'',$$

where $N_* = (\nabla g_j(x^*), j \in \tilde{I})$, $D^* = \mathrm{diag}(D_j^*, j \in \tilde{I})$, and D_j^* is defined by

$$D_j^* = \begin{cases} |g_j(x^*)|^p(|g_j(x^*) + \nabla g_j(x^*)^T d^* - \gamma_j \eta_* z_*| + |\eta_* z_*| + \|d^*\|), & \text{if } j \in \tilde{I}^-; \\ |\varphi(x^*) - g_j(x^*)|^p(|g_j(x^*) + \nabla g_j(x^*)^T d^* - \gamma_j \eta_* z_* - \varphi(x^*)| + |\eta_* z_*| + \|d^*\|), & \text{if } j \in \tilde{I}^+. \end{cases}$$

It is easy to know that $D_j^* \geq 0, j \in \tilde{I}$ and $D_j^* > 0$ for $j \in \tilde{I} \setminus I(x^*, 0)$. Therefore, similar to the proof of Lemma 2.2, we can conclude that V_* is nonsingular, so $\|V_k^{-1}\| \to \|V_*^{-1}\|, k \in K''$, which contradicts (3.3), thus the conclusion (iii) holds. □

Lemma 3.2. *Suppose that Assumptions A1–A3 hold, and that the sequence $\{x^k\}$ of points is generated by the proposed algorithm. Let K be an infinite index set such that $\lim_{k\in K} x^k = x^*$, $\lim_{k\in K} d^k = 0$ and $\lim_{k\in K} z_k = 0$. Then x^* is a KKT point for (1.1). Furthermore, there is an infinite index set $K' \subseteq K$ such that $\{\frac{\lambda^k}{\mu_k}, k \in K'\}$ converges to the KKT multiplier associated x^* with $\lim_{k\in K'} \mu_k = \mu_* > 0$.*

Proof. From Lemma 3.1(ii), we may assume without loss of generality that there exists an infinite index set $K' \subseteq K$ such that (3.2) holds and $\mu_k \to \mu_*$, $\lambda^k \to \lambda^*$, $k \in K'$.

Obviously, we have $L \subseteq I(x^*, 0)$ from $\lim_{k\in K} d^k = 0$ and $\lim_{k\in K} z_k = 0$, so passing to the limit $k \in K'$ and $k \to \infty$ in the formulas (2.4)–(2.8), we obtain

$$\left.\begin{aligned} &\gamma_0\mu_* + \eta_* \sum_{j\in L} \gamma_j \lambda_j^* = \gamma_0, \\ &\mu_* \nabla f(x^*) + \sum_{j\in L} \lambda_j^* \nabla g_j(x^*) = 0, \\ &\mu_* \gamma \varphi(x^*)^\sigma = 0, \quad \mu_* \ge 0, \qquad \lambda_j^* \ge 0, \quad j \in L. \end{aligned}\right\}$$

From Assumption A2 and $L \subseteq I(x^*, 0)$ as well as the relationships above, we get $\mu_* \ne 0$, that is $\mu_* > 0$, so it follows $\varphi(x^*) = 0$. Thus, x^* is a KKT point for (1.1) with KKT multiplier λ^*/μ_*. $\square$

Lemma 3.3. *If Assumptions A1–A3 hold, and $\lim_{k\in K} \eta_k = \eta_* > 0$, then $\lim_{k\in K} d^k = 0$ and $\lim_{k\in K} z_k = 0$.*

Proof. Firstly, we prove that $\lim_{k\in K} d^k = 0$. Suppose by contradiction that $\lim_{k\in K} d^k \ne 0$. Then there exist an infinite index subset $K' \subseteq K$ and a constant $\delta > 0$ such that $\|d^k\| \ge \delta$, $\forall k \in K'$. Thus, from Lemma 2.1 (iii) and Assumption A3, we have

$$\gamma_0 z_k \le -\frac{1}{2}(d^k)^T B_k d^k \le -\frac{1}{2} a\|d^k\|^2 \le -\frac{1}{2} a\delta^2, \qquad z_k \le -\frac{1}{2\gamma_0} a\delta^2, \quad \forall k \in K'. \tag{3.4}$$

Again, from $\lim_{k\in K} \eta_k = \eta_* > 0$, it follows that

$$\eta_k \ge \frac{\eta_*}{2}, \quad \text{for } k \in K \text{ large enough}. \tag{3.5}$$

The rest of the proof is divided into two steps as follows.

A. Show that there exists a constant $\bar t > 0$ such that the step size $t_k \ge \bar t$ for $k \in K'$.

Analyze the inequality (2.14): using Taylor expansion, from Lemma 3.1, (2.3) and (3.4), we have

$$\begin{aligned} &f(x^k + td^k + t^2\tilde d^k) - f(x^k) - \alpha t \nabla f(x^k)^T d^k - (1-\alpha)t\gamma\varphi(x^k)^\sigma \\ &\quad = f(x^k) + t\nabla f(x^k)^T d^k + o(t) - f(x^k) - \alpha t \nabla f(x^k)^T d^k - (1-\alpha)t\gamma\varphi(x^k)^\sigma \\ &\quad = (1-\alpha)t(\nabla f(x^k)^T d^k - \gamma\varphi(x^k)^\sigma) + o(t) \\ &\quad \le (1-\alpha)t\gamma_0 z_k + o(t) \le (1-\alpha)\left(-\frac{1}{2}a\delta^2\right) t + o(t). \end{aligned} \tag{3.6}$$

Thus, the inequality (2.14) holds for $k \in K'$ large enough and $t > 0$ small enough.

Analyze the inequality (2.15): for each $j \in I_k^+$, from Taylor expansion and taking into account (2.3), (3.4) and (3.5), we have for $t > 0$ sufficiently small and $k \in K'$ large enough

$$\begin{aligned} &g_j(x^k + td^k + t^2\tilde d^k) - \varphi(x^k) - \frac{t}{2}\gamma_j\eta_k z_k \\ &\quad = g_j(x^k) + t\nabla g_j(x^k)^T d^k + o(t) - \varphi(x^k) - \frac{t}{2}\gamma_j\eta_k z_k \\ &\quad \le g_j(x^k) + t\gamma_j\eta_k z_k + t\varphi(x^k) - tg_j(x^k) - \varphi(x^k) - \frac{t}{2}\gamma_j\eta_k z_k + o(t) \\ &\quad = (1-t)[g_j(x^k) - \varphi(x^k)] + \frac{t}{2}\gamma_j\eta_k z_k + o(t) \\ &\quad \le \frac{t}{2}\gamma_j\eta_k z_k + o(t) \le -\frac{\gamma_j\eta_* a\delta^2}{8\gamma_0} t + o(t) \le 0. \end{aligned}$$

For $j \in I^+(x^k) \setminus I_k^+$, it follows that $g_j(x^k) - \varphi(x^k) < -\varepsilon < 0$. So the following inequality holds for $t > 0$ small enough and $k \in K'$ large enough

$$\begin{aligned} g_j(x^k + td^k + t^2\tilde d^k) - \varphi(x^k) - \frac{t}{2}\gamma_j\eta_k z_k &= g_j(x^k) + t\nabla g_j(x^k)^T d^k + o(t) - \varphi(x^k) - \frac{t}{2}\gamma_j\eta_k z_k \\ &= [g_j(x^k) - \varphi(x^k)] + O(t) \le -\varepsilon + O(t) \le 0. \end{aligned}$$

Hence the inequality (2.15) holds for $t > 0$ small enough and $k \in K'$ large enough.

Similar to the analysis for inequality (2.15), we can show that inequality (2.16) also holds for $k \in K'$ large enough and $t > 0$ sufficiently small.

Summarizing the analysis above, we can conclude that there exists a constant $\bar{t} > 0$ such that $t_k \geq \bar{t}$ for all $k \in K'$.

B. Use $t_k \geq \bar{t} > 0$ to bring a contradiction. Two cases are considered as follows.

Case A: There exists an iteration index s such that $x^s \in X$ and

$$\varphi(x^k) = 0, \quad x^k \in X, \qquad I^+(x^k) \equiv \emptyset, \qquad I^-(x^k) \equiv I, \quad \forall k \geq s.$$

In view of the inequalities (2.14) and (2.3) as well as (3.4), we have that

$$f(x^{k+1}) - f(x^k) \leq \alpha t_k \nabla f(x^k)^T d^k \leq \alpha \gamma_0 z_k t_k, \quad \forall k \geq s, \tag{3.7}$$

$$f(x^{k+1}) - f(x^k) \leq \alpha t_k \nabla f(x^k)^T d^k \leq \alpha \gamma_0 z_k t_k \leq -\frac{1}{2}\alpha a \delta^2 \bar{t}, \quad \forall k \in K', k \geq s. \tag{3.8}$$

Thus, the whole sequence $\{f(x^k)\}_{k\geq s}$ is decreasing. Further taking into account $\lim_{k\in K'} f(x^k) = f(x^*)$, we know that $\lim_{k\to\infty} f(x^k) = f(x^*)$. So passing to the limit $k \in K'$ and $k \to \infty$ in (3.8), we have that $-\frac{1}{2}\alpha a \delta^2 \bar{t} \geq 0$, which is a contradiction.

Case B: In the case where $x^k \notin X, I^+(x^k) \neq \phi, \forall k$, from (2.15) we have that

$$\varphi(x^{k+1}) \leq \varphi(x^k) + \frac{t_k}{2}\gamma_{\min}\eta_k z_k,$$

where $\gamma_{\min} = \min\{\gamma_j, j \in I\}$. So the sequence $\{\varphi(x^k)\}_{k\geq 0}$ is decreasing and therefore $\lim_{k\to\infty} \varphi(x^k) = \varphi(x^*)$ follows from $\lim_{k\in K'} \varphi(x^k) = \varphi(x^*)$. Furthermore,

$$0 = \lim_{k\in K'}(\varphi(x^{k+1}) - \varphi(x^k)) \leq \lim_{k\in K'}\left(\frac{t_k}{2}\gamma_{\min}\eta_k z_k\right) \leq -\frac{\bar{t} a \eta_*}{8\gamma_0}\gamma_{\min}\delta^2.$$

This inequality contradicts the fact that $\bar{t} > 0, a > 0, \eta_* > 0, \gamma_0 > 0, \gamma_{\min} > 0$ and $\delta > 0$.

From the discussion above, we bring a contradiction either Case A or Case B takes place, so $\lim_{k\in K} d^k = 0$. Secondly, we show $\lim_{k\in K} z_k = 0$. If Case A happens, then from $\nabla f(x^k)^T d^k \leq \gamma_0 z_k \leq 0$ and $\lim_{k\in K} d^k = 0$ as well as $\lim_{k\in K} x^k = x^*$, we know that the claim holds. If Case B takes place, then from $0 \geq z_k \geq \frac{1}{\gamma_j \eta_k}\nabla g_j(x^k)^T d^k$ for $j \in I^+(x^k, 0) \neq \emptyset$ and $\lim_{k\in K} d^k = 0$, we also know that $\lim_{k\in K} z_k = 0$ holds. Thus, the whole proof is completed. □

Theorem 3.1. *If Assumptions A1–A3 are satisfied, then the proposed algorithm either stops at a KKT point for problem (1.1) after a finite number of iterations, or generates an infinite sequence $\{x^k\}$ of points such that each accumulation x^* of $\{x^k\}$ is a KKT point for problem (1.1).*

Proof. If the algorithm stops at the k-th iteration, then from Step 1 and Lemma 2.1, it follows that the current iteration point x^k is a KKT point for (1.1).

Now suppose that an infinite sequence $\{x^k\}$ of iterates is generated by the algorithm and that x^* is a given accumulation of $\{x^k\}$. From Lemma 3.1, we can assume without loss of generality that there exists an infinite index set K, such that the relationship (3.2) holds and

$$\mu_k \to \mu_*, \qquad \lambda^k \to \lambda^*, \quad k \in K.$$

The cases $\eta_* > 0$ and $\eta_* = 0$ are discussed, respectively.

In the former case, i.e., $\eta_* > 0$, from Lemmas 3.3 and 3.2 we know that x^* is a KKT point for (1.1).

In the latter case, from the definition of η_k in Step 4, we have

$$\eta_k = \|d^{k-1}\|^{\xi} + |z_{k-1}|^{\zeta} + \varphi(x^{k-1}) \to 0, \quad k \in K,$$

that is

$$\lim_{k\in K} d^{k-1} = 0, \qquad \lim_{k\in K} z_{k-1} = 0, \qquad \lim_{k\in K} \varphi(x^{k-1}) = 0.$$

Furthermore, from Steps 2 and 4, it follows that

$$\lim_{k\in K} \|x^k - x^{k-1}\| \leq \lim_{k\in K}(t_{k-1}\|d^{k-1}\| + t_{k-1}^2\|\tilde{d}^{k-1}\|) \leq \lim_{k\in K}(2\|d^{k-1}\| + \varphi(x^{k-1})^{\sigma'}) = 0.$$

This together with $\lim_{k\in K} x^k = x^*$ shows that $\lim_{k\in K} x^{k-1} = x^*$. Summarizing the discussion above, and letting $\bar{K} = \{k-1, k \in K\}$, we have

$$\lim_{k\in \bar{K}} x^k = x^*, \qquad \lim_{k\in \bar{K}} d^k = 0, \qquad \lim_{k\in \bar{K}} z_k = 0$$

which together with Lemma 3.2 shows that x^* is a KKT point for (1.1). □

4. Superlinear convergence analysis

In this section, we first discuss the strong convergence of the proposed algorithm, and then prove that the algorithm is superlinearly convergent under the LICQ at the limit point x^* but without assuming the strict complementarity. We make the following additional hypothesis.

Assumption A4. (i) The functions $f(x), g_j(x)(j \in I)$ are all twice continuously differentiable.

(ii) The sequence $\{x^k\}$ of points is bounded and possesses an accumulation point x^* such that the LICQ holds at x^* (so from Theorem 3.1 we know that x^* is a KKT point with a unique KKT multiplier u^* for (1.1)), and the KKT pair (x^*, u^*) satisfies the strong second-order sufficient conditions, i.e.,

$$d^T \nabla^2_{xx} L(x^*, u^*) d > 0, \quad \forall d \in \Omega \stackrel{\text{def}}{=} \{d \in R^n : d \neq 0, \nabla g_{I_*^+}(x^*)^T d = 0\},$$

where

$$L(x, u) = f(x) + \sum_{j \in I} u_j g_j(x), \qquad I_*^+ = \{j \in I : u_j^* > 0\}.$$

Under the stated assumptions, we show that the algorithm is strongly convergent.

Theorem 4.1. *If Assumptions A2–A4 are all satisfied, then*

$$\lim_{k\to\infty} \varphi(x^k) = 0, \qquad \lim_{k\to\infty} z_k = 0, \qquad \lim_{k\to\infty} d^k = 0, \qquad \lim_{k\to\infty} \|x^{k+1} - x^k\| = 0$$

and the proposed algorithm is strongly convergent, i.e., $\lim_{k\to\infty} x^k = x^*$.

Proof. In view of Theorem 3.1, we know that the given point x^* is a KKT point for (1.1), so $\varphi(x^*) = 0$. Furthermore, by the monotone property of $\{\varphi(x^k)\}$, we have

$$\lim_{k\to\infty} \varphi(x^k) = \varphi(x^*) = 0, \qquad I_* \triangleq I(x^*, 0) = \{j \in I : g_j(x^*) = 0\}. \tag{4.1}$$

Now we prove that $\lim_{k\to\infty} d^k = 0$. Suppose by contradiction that there exist an infinite index set K and a constant $\delta > 0$ such that $\|d^k\| \geq \delta$ for all $k \in K$. From Lemma 3.1 and the boundedness of $\{x^k\}$, we can assume that there exists an infinite index set $K' \subseteq K$ such that

$$\lim_{k\in K'} x^k = \bar{x}^*, \qquad \lim_{k\in K'} d^k = d^* \neq 0, \qquad \lim_{k\in K'} \eta_k = \eta_*,$$
$$I_k^+ \equiv \tilde{I}^+, \qquad I_k^- \equiv \tilde{I}^-, \quad \forall k \in K', \qquad \varphi(\bar{x}^*) = 0, \qquad I(\bar{x}^*, 0) \subseteq \tilde{I}^+ \cup \tilde{I}^-.$$

Thus, from $\gamma_0 z_k \leq -\frac{1}{2}(d^k)^T B_k d^k$ and Assumption A3, we have

$$z_* \stackrel{\text{def}}{=} \lim_{k\in K'} z_k \leq \lim_{k\in K'} \left(-\frac{1}{2\gamma_0} (d^k)^T B_k d^k \right) \leq -\frac{1}{2\gamma_0} a \|d^*\|^2 < 0.$$

If $j \in I(\bar{x}^*, 0)$, then we have $g_j(\bar{x}^*) = 0$ and $\lim_{k\in K'} g_j(x^k) = g_j(\bar{x}^*) = 0$. Combining with the constraints of (2.3) and $\lim_{k\to\infty} \varphi(x^k) = 0$, we have

$$\nabla f(\bar{x}^*)^T d^* \leq \gamma_0 z_* < 0, \qquad \nabla g_j(\bar{x}^*)^T d^* \leq \gamma_j \eta_* z_* \leq 0, \quad j \in I(\bar{x}^*, 0). \tag{4.2}$$

Letting $\bar{u}^*$ be the multiplier vector corresponding to $\bar{x}^*$, we obtain

$$\nabla f(\bar{x}^*) + \sum_{j \in I(\bar{x}^*, 0)} \bar{u}_j^* \nabla g_j(\bar{x}^*) = 0, \quad \bar{u}_j^* \geq 0, j \in I(\bar{x}^*, 0), \tag{4.3}$$

which contradicts (4.2). Therefore, $\lim_{k\to\infty} d^k = 0$.

The claim $\lim_{k\to\infty} z_k = 0$ follows immediately from $\nabla f(x^k)^T d^k - \gamma \varphi(x^k)^\sigma \leq \gamma_0 z_k \leq 0$. Thus,

$$\lim_{k\to\infty} \|x^{k+1} - x^k\| \leq \lim_{k\to\infty} (t_k \|d^k\| + t_k^2 \|\tilde{d}^k\|) \leq \lim_{k\to\infty} (2\|d^k\| + \varphi(x^k)^{\sigma'}) = 0.$$

Finally, we can conclude that the given accumulation point x^* is an isolated KKT point for (1.1) under Assumption A4. Therefore, x^* is an isolated accumulation point of $\{x^k\}$ from Theorem 3.1, and this together with $\lim_{k\to\infty} \|x^{k+1} - x^k\| = 0$ shows that $\lim_{k\to\infty} x^k = x^*$. (The proof can be found in [18].) □

Lemma 4.1. *Suppose that Assumptions A2–A4 are all satisfied. Then*

(i) *for k large enough, the coefficient matrix V_k in the system of linear equations (2.9) is nonsingular, and there exists a constant $\bar{c}$ such that $\|V_k^{-1}\| \le \bar{c}$;*

(ii) $\lim_{k\to\infty} \eta_k = 0$, $\lim_{k\to\infty} \mu_k = 1$, $\lim_{k\to\infty} \lambda^k = u^*$;

(iii) *for k large enough, the relationship $I_*^+ \subseteq L_k \subseteq I_* \subseteq I_k$ holds, where the set I_* is defined by (4.1).*

Proof. (i) Based on Theorem 4.1, in view of the fact that the subsets I_k all are subsets of the fixed and finite set I, it is sufficient to show that any accumulation point V_* of $\{V_k\}$ is nonsingular. In view of (2.10)–(2.12), we can conclude that there is an infinite index set K such that

$$V_k \to V_* = \begin{pmatrix} B_* & N_* \\ N_*^T & -D^* \end{pmatrix}, \qquad B_k \to B_*, \qquad I_k^- \equiv \tilde{I}^-, \qquad I_k^+ \equiv \tilde{I}^+, \qquad I_k \equiv \tilde{I}^+ \cup \tilde{I}^- \triangleq \tilde{I}, \quad k \in K,$$

where $N_* = (\nabla g_j(x^*), j \in \tilde{I})$, $D^* = \mathrm{diag}(|g_j(x^*)|^{p+1}, j \in \tilde{I})$.

So similar to the proof of Lemma 2.2, it follows that V_* is nonsingular.

(ii) It is obvious that $\lim_{k\to\infty} \eta_k = 0$ from $\lim_{k\to\infty} d^k = 0$, $\lim_{k\to\infty} z_k = 0$ and $\lim_{k\to\infty} \varphi(x^k) = 0$ as well as the definition of η_k in Step 4. According to Lemma 3.1 (ii), it follows that the sequences $\{\mu_k\}$ and $\{\lambda^k\}$ are bounded. Therefore, in view of (2.4) and $\lim_{k\to\infty} \eta_k = 0$, we have $\lim_{k\to\infty} \mu_k = 1$. Furthermore, combining Lemma 3.2, Theorem 4.1 with the uniqueness of the KKT multiplier vector u^*, we can conclude that $\lim_{k\to\infty} \lambda^k = u^*$.

(iii) Since

$$\lim_{k\to\infty} x^k = x^*, \qquad \lim_{k\to\infty} \varphi(x^k) = 0, \qquad \lim_{k\to\infty} \eta_k = \lim_{k\to\infty} z_k = 0, \qquad \lim_{k\to\infty} d^k = 0,$$

we have $L_k \subseteq I_*$. Then for $j \in I_*^+$, i.e., $u_j^* > 0$, we obtain that $\lambda_j^k > 0$ from (ii), so $j \in L_k$ and $I_*^+ \subseteq L_k$ for k sufficiently large. In addition, for $j \in I_*$ and k sufficiently large, we have $-\varepsilon \le g_j(x^k) - \varphi(x^k) \le 0$ or $-\varepsilon \le g_j(x^k) \le 0$, so $j \in I_k$. □

The following relations will be fundamental to the rest of analysis.

Lemma 4.2. *Under Assumptions A2–A4, the following relations hold*

$$|z_k| = O(\|d^k\|) + O(\varphi(x^k)^\sigma), \tag{4.4}$$

$$\|\tilde{d}^k\| = O(\|d^k\|^2) + O(\varphi(x^k)^\sigma) + O(|\eta_k z_k|), \qquad \|h^k\| = O(\|d^k\|^2) + O(\varphi(x^k)^\sigma) + O(|\eta_k z_k|), \tag{4.5}$$

$$\|\tilde{d}^k\|^2 = O(\|d^k\|^4) + o(\varphi(x^k)^\sigma) + o(|\eta_k z_k|\|d^k\|), \tag{4.6}$$

$$\|d^k\| \cdot \|\tilde{d}^k\| = O(\|d^k\|^3) + o(\varphi(x^k)^\sigma) + O(|\eta_k z_k|\|d^k\|), \tag{4.7}$$

where $(\tilde{d}^k, h^k)$ is generated by the system of equations (2.9). Furthermore, the correction direction $\tilde{d}^k$ in the algorithm is always obtained from the solution of (2.9) for k large enough.

Proof. By the first constraint of (2.3), we have $\frac{1}{\gamma_0}(\nabla f(x^k)^T d^k - \gamma\varphi(x^k)^\sigma) \le z_k \le 0$. Hence

$$|z_k| \le \left|\frac{1}{\gamma_0}(\nabla f(x^k)^T d^k - \gamma\varphi(x^k)^\sigma)\right| \le \frac{1}{\gamma_0}\|\nabla f(x^k)\|\|d^k\| + \frac{\gamma}{\gamma_0}\varphi(x^k)^\sigma,$$

which shows that the relation (4.4) holds. Furthermore, in view of formulas (2.9)–(2.12), $\tau \in (2, 3)$ and Lemma 4.1 (i) as well as Taylor expansion, we have

$$\begin{aligned}
\left\|\begin{pmatrix} \tilde{d}^k \\ h^k \end{pmatrix}\right\| &= O(\max\{\|d^k\|^\tau, |\eta_k^\nu z_k|\|d^k\|\} + \|\tilde{g}^k\|) \\
&= O(\max\{\|d^k\|^\tau, |\eta_k^\nu z_k|\|d^k\|\} + O(\|d^k\|^2) + O(|\eta_k z_k|) + O(\varphi(x^k)^\sigma)) \\
&= O(\max\{\|d^k\|^\tau, \eta_k^\nu(O(\|d^k\|) + O(\varphi(x^k)^\sigma))\|d^k\|\}) + O(\|d^k\|^2) + O(|\eta_k z_k|) + O(\varphi(x^k)^\sigma) \\
&= O(\max\{\|d^k\|^\tau, o(\|d^k\|^2) + o(\varphi(x^k)^\sigma)\}) + O(\|d^k\|^2) + O(|\eta_k z_k|) + O(\varphi(x^k)^\sigma) \\
&= O(\|d^k\|^2) + O(|\eta_k z_k|) + O(\varphi(x^k)^\sigma).
\end{aligned}$$

Therefore, the relation (4.5) holds. The rest two relations (4.6) and (4.7) are easy to be obtained by (4.5). Finally, from (4.4) and (4.5) and $\sigma' \in (0, \sigma)$, we know that the vector $\tilde{d}^k$ computed by (2.9)–(2.12) satisfies

$$\begin{aligned}
\|\tilde{d}^k\| &= O(\|d^k\|^2) + O(\varphi(x^k)^\sigma) + O(\eta_k\|d^k\|) + O(\eta_k\varphi(x^k)^\sigma) \\
&= o(\|d^k\|) + O(\varphi(x^k)^\sigma) \le \|d^k\| + \varphi(x^k)^{\sigma'}.
\end{aligned}$$

Thus, taking into account Lemma 4.1 (i) and Step 2, we can conclude that the correction direction $\tilde{d}^k$ is always generated by the system of linear equations (2.9) for k large enough. □

Theorem 4.2. *Suppose that Assumptions A2–A4 hold. Then $g_j(x^k + d^k + \tilde{d}^k) \le 0, j \in I$ hold for k large enough, and therefore $\varphi(x^{k+1}) = 0$, i.e., the iterates get into the feasible set after a finite number of iterations.*

Proof. In view of Step 3 of the algorithm, it is sufficient to show that $g_j(x^k + d^k + \tilde{d}^k) \le 0, j \in I$ for k large enough. We consider two cases as follows.

First, for $j \notin I_*$, i.e., $g_j(x^*) < 0$, taking into account $(x^k, d^k, \tilde{d}^k) \to (x^*, 0, 0)$, we can conclude by continuity that $g_j(x^k + d^k + \tilde{d}^k) \le 0$ holds for $j \notin I_*$ and k large enough.

Second, for $j \in I_*$, i.e., $g_j(x^*) = 0$. From Lemma 4.1 (iii), we obtain that $j \in I_k$. For simplicity, we denote

$$w_j^k = \begin{cases} g_j(x^k) + \nabla g_j(x^k)^T d^k - \gamma_j \eta_k z_k, & \text{if } j \in I_k^-; \\ g_j(x^k) + \nabla g_j(x^k)^T d^k - \gamma_j \eta_k z_k - \varphi(x^k), & \text{if } j \in I_k^+. \end{cases}$$

Hence from (2.12), we have

$$|D_j^k| = o(|w_j^k|) + o(|\eta_k z_k|) + o(\|d^k\|). \tag{4.8}$$

By Taylor expansion, it follows that

$$\begin{aligned} g_j(x^k + d^k + \tilde{d}^k) &= g_j(x^k + d^k) + \nabla g_j(x^k + d^k)^T \tilde{d}^k + O(\|\tilde{d}^k\|^2) \\ &= g_j(x^k + d^k) + \nabla g_j(x^k)^T \tilde{d}^k + O(\|d^k\| \cdot \|\tilde{d}^k\|) + O(\|\tilde{d}^k\|^2). \end{aligned}$$

According to (4.6) and (4.7), we have

$$O(\|\tilde{d}^k\| \cdot \|d^k\|) + O(\|\tilde{d}^k\|^2) = O(\|d^k\|^3) + O(|\eta_k z_k| \|d^k\|) + o(\varphi(x^k)^\sigma). \tag{4.9}$$

So

$$g_j(x^k + d^k + \tilde{d}^k) = g_j(x^k + d^k) + \nabla g_j(x^k)^T \tilde{d}^k + O(\|d^k\|^3) + O(|\eta_k z_k| \|d^k\|) + o(\varphi(x^k)^\sigma). \tag{4.10}$$

On the other hand, from (2.9) we obtain

$$\nabla g_j(x^k)^T \tilde{d}^k = -\max\{\|d^k\|^\tau, |\eta_k^\nu z_k| \|d^k\|\} - \tilde{g}_j^k + D_j^k h_j^k. \tag{4.11}$$

Furthermore, from (2.10)–(2.12), (4.4), (4.5) and (4.8), we have

$$\begin{aligned} g_j(x^k + d^k) + \nabla g_j(x^k)^T \tilde{d}^k &= -\max\{\|d^k\|^\tau, |\eta_k^\nu z_k| \|d^k\|\} - \varphi(x^k)^\sigma + w_j^k + o(|w_j^k|) \\ &\quad + o(\|d^k\|^3) + o(\varphi(x^k)^\sigma) + o(|\eta_k z_k| \|d^k\|). \end{aligned} \tag{4.12}$$

Substituting (4.12) into (4.10) and in view of the constraints of (2.3) as well as $\nu \in (0, 1)$, we get

$$\begin{aligned} g_j(x^k + d^k + \tilde{d}^k) &= -\max\{\|d^k\|^\tau, |\eta_k^\nu z_k| \|d^k\|\} - \varphi(x^k)^\sigma + w_j^k + o(|w_j^k|) \\ &\quad + O(\|d^k\|^3) + o(\varphi(x^k)^\sigma) + O(|\eta_k z_k| \|d^k\|) \le 0 \end{aligned} \tag{4.13}$$

holds for $j \in I_*$ and k sufficiently large.

Therefore, in the case of $\varphi(x^k) > 0$, from Step 3, we have $t_k = 1$ and $x^{k+1} = x^k + d^k + \tilde{d}^k$, thus $\varphi(x^{k+1}) = 0$ holds for k large enough. The proof is complete. □

To ensure the unit step can be accepted by the arc search for k large enough and meanwhile remove the strict complementarity assumption, we need the following assumption.

Assumption A5. Suppose that the KKT pair (x^*, u^*) and the matrix B_k satisfy

$$\|(\nabla_{xx}^2 L(x^*, u^*) - B_k) d^k\| = o(\|d^k\|).$$

Remark 3. According to Theorem 4.1 and Lemma 4.1, it is easy to know that Assumption A5 is equivalent to

$$\|(\nabla_{xx}^2 L(x^k, \lambda^k / \mu_k) - B_k) d^k\| = o(\|d^k\|).$$

Theorem 4.3. *Suppose that Assumptions A2–A5 hold. Then the unit step is accepted by the arc search (2.14)–(2.16), i.e., $t_k \equiv 1$, for k sufficiently large.*

Proof. By Theorem 4.2, we have that $\varphi(x^k) \equiv 0$ and $I^+(x^k) \equiv \emptyset$ for k large enough, moreover, (2.15) and (2.16) hold for $t = 1$ and k large enough. In fact, the inequality (2.15) vanishes for sufficiently large k. So we only need to check that (2.14) holds for $t = 1$ and k large enough.

Firstly, from (2.9)–(2.12) and Lemma 4.2 as well as Theorem 4.2, we have that, for k large enough

$$\begin{aligned}
&x^k \in X, \quad \varphi(x^k) = 0, \qquad I_k^+(x^k) = \emptyset, \qquad I_k^-(x^k) \equiv I,\\
&\tilde{g}_j^k = g_j(x^k + d^k) - g_j(x^k) - \nabla g_j(x^k)^T d^k + \gamma_j \eta_k z_k, \quad \forall j \in I_k,\\
&D_j^k = |g_j(x^k)|^p (|g_j(x^k) + \nabla g_j(x^k)^T d^k - \gamma_j \eta_k z_k| + |\eta_k z_k| + \|d^k\|), \quad j \in I_k,\\
&|z_k| = O(\|d^k\|), \qquad \|\tilde{d}^k\| = o(\|d^k\|), \qquad L_k = \{j \in I_k : g_j(x^k) + \nabla g_j(x^k)^T d^k = \gamma_j \eta_k z_k\}.
\end{aligned}$$

Let $\omega_k \overset{\text{def}}{=} f(x^k + d^k + \tilde{d}^k) - f(x^k) - \alpha \nabla f(x^k)^T d^k$. Then it is sufficient to show that $\omega_k \le 0$. In view of Taylor expansion and $\|\tilde{d}^k\| = o(\|d^k\|)$, we have

$$\omega_k = \nabla f(x^k)^T (d^k + \tilde{d}^k) + \frac{1}{2}(d^k)^T \nabla^2 f(x^k) d^k - \alpha \nabla f(x^k)^T d^k + o(\|d^k\|^2). \tag{4.14}$$

According to (2.5), we get

$$\nabla f(x^k) = -\frac{1}{\mu_k} B_k d^k - \frac{1}{\mu_k} \sum_{j \in L_k} \lambda_j^k \nabla g_j(x^k).$$

This together with the definition of L_k and $\|\tilde{d}^k\| = o(\|d^k\|)$ shows that

$$\begin{aligned}
\nabla f(x^k)^T d^k &= -\frac{1}{\mu_k}(d^k)^T B_k d^k - \frac{1}{\mu_k} \sum_{j \in L_k} \lambda_j^k \nabla g_j(x^k)^T d^k\\
&= -\frac{1}{\mu_k}(d^k)^T B_k d^k + \frac{1}{\mu_k} \sum_{j \in L_k} \lambda_j^k g_j(x^k) - \frac{1}{\mu_k} \sum_{j \in L_k} \gamma_j \lambda_j^k \eta_k z_k,
\end{aligned} \tag{4.15}$$

and

$$\nabla f(x^k)^T \tilde{d}^k = -\frac{1}{\mu_k} \sum_{j \in L_k} \lambda_j^k \nabla g_j(x^k)^T \tilde{d}^k + o(\|d^k\|^2).$$

Therefore,

$$\nabla f(x^k)^T (d^k + \tilde{d}^k) = -\frac{1}{\mu_k}(d^k)^T B_k d^k - \frac{1}{\mu_k} \sum_{j \in L_k} \lambda_j^k \nabla g_j(x^k)^T (d^k + \tilde{d}^k) + o(\|d^k\|^2). \tag{4.16}$$

Secondly, for $j \in L_k \subseteq I_* = I_* \cap I^-(x^k)$, from (4.13) and Taylor expansion, it is easy to prove that $w_j^k = 0$, $|z_k| = O(\|d^k\|)$, and

$$o(\|d^k\|^2) = g_j(x^k + d^k + \tilde{d}^k) = g_j(x^k) + \nabla g_j(x^k)^T (d^k + \tilde{d}^k) + \frac{1}{2}(d^k)^T \nabla^2 g_j(x^k) d^k + o(\|d^k\|^2).$$

Hence we have

$$-\frac{1}{\mu_k} \sum_{j \in L_k} \lambda_j^k \nabla g_j(x^k)^T (d^k + \tilde{d}^k) = \frac{1}{\mu_k} \sum_{j \in L_k} \lambda_j^k g_j(x^k) + \frac{1}{2}(d^k)^T \left(\sum_{j \in L_k} \frac{\lambda_j^k}{\mu_k} \nabla^2 g_j(x^k) \right) d^k + o(\|d^k\|^2).$$

Substituting the above equality into (4.16), and further substituting the associated result and (4.15) into (4.14), we obtain

$$\begin{aligned}
\omega_k = {}& \left[\frac{1}{\mu_k}(\alpha - 1) + \frac{1}{2} \right] (d^k)^T B_k d^k + \frac{1}{2}(d^k)^T \left(\nabla^2 f(x^k) + \sum_{j \in L_k} \frac{\lambda_j^k}{\mu_k} \nabla^2 g_j(x^k) - B_k \right) d^k\\
&+ \frac{1-\alpha}{\mu_k} \sum_{j \in L_k} \lambda_j^k g_j(x^k) + \frac{\alpha}{\mu_k} \sum_{j \in L_k} \gamma_j \lambda_j^k \eta_k z_k + o(\|d^k\|^2).
\end{aligned}$$

So, in view of $\mu_k \to 1$, $\alpha \in (0, \frac{1}{2})$ and Assumptions A3 and A5, for k large enough, we have

$$\begin{aligned}
\omega_k &\le \left[\frac{1}{\mu_k}(\alpha - 1) + \frac{1}{2} \right] a\|d^k\|^2 + \frac{1-\alpha}{\mu_k} \sum_{j \in L_k} \lambda_j^k g_j(x^k) + \frac{\alpha}{\mu_k} \sum_{j \in L_k} \gamma_j \lambda_j^k \eta_k z_k + o(\|d^k\|^2)\\
&\le \left[\frac{1}{\mu_k}(\alpha - 1) + \frac{1}{2} \right] a\|d^k\|^2 + o(\|d^k\|^2) \le 0.
\end{aligned}$$

Thus, the inequality (2.14) holds for $t = 1$ and k large enough. The whole proof is completed. □

Based on Theorems 4.2, 4.3 and $\|\tilde{d}^k\| = o(\|d^k\|)$, similar to the proof of Theorem 4.3 in [17], we can prove the superlinear convergence of the algorithm.

Theorem 4.4. *Suppose that Assumptions A2–A5 hold. Then the proposed algorithm is superlinearly convergent, i.e., the sequence* $\{x^k\}$ *satisfies* $\|x^{k+1} - x^*\| = o(\|x^k - x^*\|)$.

5. Numerical experiments

In order to illustrate the computational effectiveness of the proposed algorithm, we test some typical Hock and Schittkowski's problems which are taken from [19,20] and also collected in the widely used testing environment CUTEr [21]. In addition, we test a large-scale problem `Svanberg` (in different dimensions) which is also taken from CUTEr [21]. The algorithm was implemented by using Matlab 7.1 on Windows XP platform, and on a PC with 1700 MHz CPU. The approximate Hessian matrix B_k is updated by the damped BFGS formula from Powell [22], and B_0 is the identity matrix. During the numerical experiments, we set

$$\begin{aligned}&\gamma_0 = 2.0, \quad \gamma_j = 1, \quad j = 1, 2, \ldots, m, \quad p = 1, \quad \eta_0 = 0.2, \quad \nu = 0.55, \quad \tau = 2.25,\\ &\beta = 0.58, \quad \alpha = 0.25, \quad \sigma = 0.7, \quad \sigma' = 0.1, \quad \gamma = 5.5, \quad c_0 = 0.2, \quad \xi = \zeta = 0.8.\end{aligned}$$

The condition $\|d^k\| \le \epsilon$ is set to be the stopping criterion.

The computational results are listed in four tables whose columns have the following meanings:

Prob: the problem number as given in [19,20];

SNQP: our algorithm;

ALGO: the algorithm in [17];

RFSQP: the algorithm in [13];

FSLE: the algorithm in [23];

n: the number of variables;

m: the number of constraints;

Ni: the number of iterations;

INi: the number of iterations before the iteration point enters the feasible set X;

Nf: the number of objective function evaluations;

Ng: the number of nonlinear constraint function evaluations for SNQP, ALGO and RFSQP, but for FSLE it denotes the number of function evaluations for $g = (g_1, \ldots, g_m)^T$;

$f(x^*)$: the final objective value.

From the above four tables, we see that our algorithm succeeded in solving all the test problems.

In Table 1, we report some Hock and Schittkowski's problems [13,17] for feasible initial points which are given in [19,20]. The results are not listed for algorithms `RFSQP` and `FSLE` for those problems that are not reported in [13,23]. From the view point of the numbers for Ni, Nf and Ng, the results in Table 1 show that `SNQP` is competitive with `RFSQP`, but is better than `ALGO` and `FSLE` for many problems. From Table 2, we see that `SNQP` is slightly better than `ALGO`, and is obviously better than `FSLE`.

In Tables 3 and 4, the numerical results for some infeasible starting points are reported, and we used $\|d^k\| \le 10^{-5}$ as the stopping criterion. Note that the problems with triangles ▲ mean that the tentative search in Step 3 is successful in a certain iteration, which make the iteration points to get into the feasible set earlier. From Table 3, we see that `SNQP` performs better than `ALGO` for the same starting points, in particular, **INi** of `SNQP` is typically low compared to `ALGO` for most problems. Finally, Table 4 shows that our algorithm is successful for all the cases, and only one iteration is needed to generate a feasible point.

6. Concluding remarks

In this paper, we proposed a new SQP algorithm based on the ideas of a norm-relaxed SQP method and a strongly sub-feasible direction method for solving inequality constrained optimization. Starting from an infeasible point, the algorithm can always generate feasible points after a finite number of earlier iterations. The master direction is generated by solving a QP subproblem, the high-order correction direction is obtained by solving a system of linear equations, and both of them consist of the constraints/gradients corresponding to the estimate of the active set. The global convergence is obtained under the MFCQ and the superlinear convergence is proved without the strict complementarity. Numerical results show that the proposed algorithm is promising.

As a further work, the techniques introduced in this paper can be extended to solve general constrained optimization problems and minimax problems. In addition, testing experience shows that it is not suitable for too small value of ε, so a general principle for choosing the parameter ε should be further studied.

Table 1
Numerical results for feasible initial point

Prob	n, m	Method	Ni	Nf	Ng	$f(x^*)$	ε	ϵ
012	2, 1	SNQP	7	7	27	−3.0000000E+01	1	1.0E−06
		ALGO	7	8	18	−3.0000000E+01		1.0E−06
		RFSQP	7	7	14	−3.0000000E+01		1.0E−06
		FSLE	7	24	28	−3.0000000E+01		
029	3, 1	SNQP	11	15	42	−2.2627417E+01	2	1.0E−05
		ALGO	9	12	21	−2.2627417E+01		1.0E−05
		RFSQP	10	11	20	−2.2627417E+01		1.0E−05
		FSLE	9	28	34	−2.2627417E+01		
030	3, 7	SNQP	8	9	16	1.0000000E+00	14	1.0E−07
		ALGO	13	15	28	1.0000000E+00		1.0E−07
		RFSQP	18	18	35	1.0000000E+00		1.0E−07
031	3, 7	SNQP	15	21	39	6.0000000E+00	1	1.0E−05
		ALGO	10	15	26	6.0000000E+00		1.0E−05
		RFSQP	8	9	36	6.0000000E+00		1.0E−05
033	3, 6	SNQP	36	44	219	−4.5857864E+00	5	1.0E−08
		ALGO	23	95	237	−4.5857864E+00		1.0E−08
		RFSQP	4	4	11	−4.0000000E+00		1.0E−08
035	3, 4	SNQP	6	6	0	1.1111111E−01	2	1.0E−06
		ALGO	6	11	0	1.1111121E−01		1.0E−06
		FSLE	7	13	19	1.11111E−01		
038	4, 8	SNQP	45	76	0	1.08914675E−06	13	1.0E−06
		ALGO	37	104	0	5.01880096E−07		1.0E−06
		FSLE	49	91	91	5.128073E−11		
043	4, 3	SNQP	12	12	77	−4.3999999E+01	5	1.0E−05
		ALGO	10	13	75	−4.4000000E+01		1.0E−05
		RFSQP	9	9	51	−4.4000000E+01		1.0E−05
		FSLE	12	36	45	−4.4000000E+01		
066	3, 8	SNQP	11	11	167	5.1816327E−01	7	1.0E−08
		ALGO	14	26	84	5.1816327E−01		1.0E−08
		RFSQP	8	8	30	5.1816327E−01		1.0E−08
093	6, 2	SNQP	16	16	743	1.3507594E+02	1	1.0E−05
		ALGO	35	138	612	1.3507594E+02		1.0E−05
		RFSQP	12	13	54	1.3507594E+02		1.0E−05
		FSLE	18	51	69	1.3507596E+02		
113	10, 8	SNQP	30	30	275	2.4306211E+01	11	1.0E−03
		ALGO	18	23	215	2.4306211E+01		1.0E−03
		RFSQP	12	12	120	2.4306211E+01		1.0E−03
		FSLE	21	58	79	2.4306209E+01		
264	4, 3	SNQP	13	13	93	−4.4113406E+01	6	1.0E−06
		ALGO	10	24	99	−4.4113407E+01		1.0E−06

Table 2
Numerical results of problem Svanberg for feasible initial point

Prob	n, m	Method	Initial point	Ni	Nf	Ng	ε	$f(x^*)$
Svanberg-10	10, 30	SNQP	$(0, 0, \ldots, 0)^T$	28	28	1753	5	15.731533
		ALGO		15	21	1050		15.731517
		FSLE		36	227	258		15.731517
Svanberg-30	30, 90	SNQP	$(0, 0, \ldots, 0)^T$	27	27	4975	7	49.142545
		ALGO		26	38	5670		49.142526
		FSLE		101	777	864		49.142526
Svanberg-50	50, 150	SNQP	$(0, 0, \ldots, 0)^T$	37	37	11762	7	82.581928
		ALGO		35	51	12750		82.581912
		FSLE		108	881	968		82.581912
Svanberg-80	80, 240	SNQP	$(0, 0, \ldots, 0)^T$	47	47	24100	7	132.749830
		ALGO		47	68	27360		132.749819
		FSLE		190	1666	1835		132.749819
Svanberg-100	100, 300	SNQP	$(0, 0, \ldots, 0)^T$	46	46	27880	7	166.197199
		ALGO		53	66	35400		166.197171
		FSLE		178	1628	1782		166.197171

Table 3
Numerical results for infeasible initial point

Prob	Initial point	Method	INi	Ni	Nf	Ng	$f(x^*)$	ε	ϵ
▲012	$(6, 6)^T$	SNQP	7	12	12	29	−2.9999999E+01	2	1.0E−06
		ALGO	25	28	29	57	−3.0000000E+01		1.0E−06
029	$(-4, -4, -4)^T$	SNQP	1	12	17	42	−2.2627416E+01	2	1.0E−05
		ALGO	1	11	14	27	−2.2627417E+01		1.0E−05
▲030	$(10, 10, 10)^T$	SNQP	10	16	16	53	1.0001551E+00	7	1.0E−07
		ALGO	1	19	20	39	1.0000000E+00		1.0E−07
▲031	$(2, 4, 7)^T$	SNQP	1	19	20	52	6.0000089E+00	1	1.0E−05
		ALGO	4	20	23	43	6.0000000E+00		1.0E−05
033	$(1, 4, 6)^T$	SNQP	2	21	23	116	−4.5857290E+00	6	1.0E−08
	$(2, 4, 6)^T$	ALGO	2	16	17	67	−4.5857863E+00		1.0E−08
▲035	$(1, 2, 3)^T$	SNQP	4	9	9	0	1.1111111E−01	1	1.0E−05
		ALGO	8	11	12	0	−3.4500000E+00		1.0E−05
043	$(0, 2, 2, 4)^T$	SNQP	1	11	11	69	−4.3999999E+01	3	1.0E−05
	$(-10, 2, -8, 5)^T$	ALGO	23	26	27	163	−4.4000000E+01		1.0E−05
▲076	$(1, 2, 3, 4)^T$	SNQP	2	14	14	0	−4.6818171E+00	4	1.0E−05
		ALGO	6	14	15	0	−4.6818182E+00		1.0E−05
113	$(0, 2, 9, 5, 0, 1, 9, 8, -10, 10)^T$	SNQP	9	29	29	258	2.4306211E+01	10	1.0E−03
	$(4, 10, 10, 2, 0, 11, 4, 0, 12, 10)^T$	ALGO	6	17	21	205	2.4306209E+01		1.0E−03
264	$(0, 0, 0, 10)^T$	SNQP	4	16	17	122	−4.4113405E+01	5	1.0E−06
	$(8, -5, 6, -4)^T$	ALGO	19	26	27	161	−4.3999999E+01		1.0E−06

Table 4
Numerical results of problem Svanberg for infeasible initial point

Prob	n, m	Method	Initial point	INi	Ni	Nf	Ng	ε	$f(x^*)$
▲Svanberg-10	10, 30	SNQP	$(0.6, 0.1, 0.1, 0, \dots, 0)^T$	1	25	25	1 480	7	15.731533
▲Svanberg-20	20, 60	SNQP	$(0.6, 0.1, 0.1, 0, \dots, 0)^T$	1	29	29	3 528	7	32.427948
▲Svanberg-30	30, 90	SNQP	$(0.6, 0.1, 0.1, 0, \dots, 0)^T$	1	35	35	6 827	7	49.142547
▲Svanberg-50	50, 150	SNQP	$(0.6, 0.1, 0.1, 0, \dots, 0)^T$	1	42	42	13 602	7	82.581931
▲Svanberg-80	80, 240	SNQP	$(0.6, 0.1, 0.1, 0, \dots, 0)^T$	1	43	43	21 712	7	132.749843
▲Svanberg-100	100, 300	SNQP	$(0.6, 0.1, 0.1, 0, \dots, 0)^T$	1	58	58	37 625	7	166.197200
▲Svanberg-200	200, 600	SNQP	$(0.6, 0.1, 0.1, 0, \dots, 0)^T$	1	94	94	114 486	7	333.441333

References

[1] N. Maratos, Exact penalty function algorithms for finite dimensional and control optimization problems, Ph.D. Thesis, Imperial College Sci. Tech., University of London, 1978.
[2] E.R. Panier, A.L. Tits, A superlinearly convergence feasible method for the solution of inequality constrained optimization, SIAM Journal on Control and Optimization 25 (4) (1987) 934–950.
[3] E.R. Panier, A.L. Tits, On combining feasibility, descent and superlinear convergence in equality constrained optimization, Mathematical Programming 59 (1993) 261–276.
[4] X. Xu, W. Wang, A mixed superlinearly convergent algorithm with nonmonotone search for constrained optimizations, Applied Mathematics. A Journal of Chinese Universities. Series B 15 (2) (2000) 211–219.
[5] Z.Y. Gao, F. Wu, A SQP feasible method for nonlinear constraints, Acta Mathematica Applicate Sinica 18 (4) (1995) 570–579.
[6] Z.Y. Gao, F. Wu, A superlinearly feasible method for nonlinear constraints, Acta Mathematica Sinica (Ser. A) 40 (6) (1997) 895–900.
[7] J.B. Jian, A superlinearly convergent feasible descent algorithm for nonlinear optimization, Journal of Mathematics (PRC) 15 (3) (1995) 319–326.
[8] J.B. Jian, S.J. Xue, A class of superlinearly feasible methods for nonlinearly constrained optimization, Journal of Mathematical Research and Exposition 19 (1) (1999) 135–140.
[9] J.B. Jian, K.C. Zhang, S.J. Xue, A superlinearly and quadratically convergent SQP type feasible method for constrained optimization, Applied Mathematics. A Journal of Chinese Universities. Series B 15 (3) (2000) 319–331.
[10] M.E. Cawood, M.M. Kostreva, Norm-relaxed method of feasible direction for solving the nonlinear programming problems, Journal of Optimization Theory and Applications 83 (1994) 311–320.
[11] X. Chen, M.M. Kostreva, A generalization of the norm-relaxed method of feasible directions, Applied Mathematics and Computation 102 (1999) 257–273.
[12] M.M. Kostreva, X. Chen, A superlinearly convergent method of feasible directions, Applied Mathematics and Computation 116 (2000) 231–244.
[13] C.T. Lawrence, A.L. Tits, A computationally efficient feasible sequential quadratic programming algorithm, SIAM Journal on Optimization 11 (2001) 1092–1118.
[14] Z.B. Zhu, K.C. Zhang, A new SQP method of feasible directions for nonlinear programming, Applied Mathematics and Computation 148 (2004) 121–134.
[15] J.B. Jian, Strong combined Phase I-Phase II methods of sub-feasible directions, Mathematics in Economics (A Chinese Journal) 12 (1995) 64–70.
[16] J.B. Jian, H.Y. Zheng, Q.J. Hu, C.M. Tang, A new norm-relaxed method of strongly sub-feasible direction for inequality constrained optimization, Applied Mathematics and Computation 168 (2005) 1–28.
[17] J.B. Jian, H.Y. Zheng, Q.J. Hu, C.M. Tang, A new superlinearly convergent norm-relaxed method of strongly sub-feasible direction for inequality constrained optimization, Applied Mathematics and Computation 182 (2006) 955–976.
[18] J.B. Jian, Researches on superlinearly and quadratically convergence algorithms for nonlinearly constrained optimization, Ph.D. Thesis, School of Xi'an Jiaotong University, Xi'an, China, 2000.
[19] K. Schittkowski, More Test Examples for Nonlinear Programming Codes, Springer Verlag, 1987.
[20] W. Hock, K. Schittkowski, Test Examples for Nonlinear Programming Codes, in: Lecture Notes in Economics and Mathematical systems, vol. 187, Springer-Verlag, Berlin, Heidelberg, New York, 1981.

[21] N.I.M. Gould, D. Orban, Ph.L. Toint, CUTEr (and SifDec), a Constrained and Unconstrained Testing Environment. Revisited, http://hsl.rl.ac.uk/cuter-www.
[22] M.J.D. Powell, A fast algorithm for nonlinearly constrained optimization calculations, in: Numerical Analysis, in: G.A. Watson (Ed.), Lecture Notes in Math, vol. 630, Springer, Berlin, 1978, pp. 144–157.
[23] Y.F. Yang, D.H. Li, L.Q. Qi, A feasible sequential system of linear equations method for inequality constrained optimization, SIAM Journal on Optimization 13 (2003) 1222–1244.